M4 COMPETITION: WHAT'S NEXT

DILEK ONKAL

Forecasting competitions play a significant (albeit controversial) role in forecasting history (Hyndman, 2018). M-Competitions started with ***M1*** using 111 time series (Makridakis and Hibon, 1979), and we now have the ***M4*** using 100,000 time series. The ***M4*** *Competition* took previous competitions to a different level by not just increasing the number of series and incorporating Machine Learning methods, but by acknowledging the critical challenges of forecast uncertainty and reproducibility. These neoteric extensions successfully serve the ***M4***'s stated objectives of (i) learning how to improve forecasting accuracy and (ii) benefiting academics and practitioners (Makridakis, et al, 2018a, 2018b). In addition to applauding the *M4 Competition*'s innovative dimensions and cardinal goals for both forecasters as well as users of forecasts, this commentary will suggest potential extensions for the *M-Competitions* to come.

Explicitly recognizing forecast uncertainty, the ***M4*** asked for and evaluated prediction intervals in addition to point forecasts. Communication of the uncertainties embedded in forecasts is critical to users of such forecasts and prediction intervals portray this to a certain extent. A more detailed profile of such uncertainty would be provided by forecast distributions, and I hope future competitions will take this challenge and include comparative evaluations of forecast distribution accuracy.

The added emphasis on reproducibility is another commendable novelty of this competition. This feature sets a new benchmark for future competitions and counteracts any criticisms on plausibility of findings via providing an open platform enabling replications.

Also, given the prevalence of big data and digital torrents faced by many sectors, other promising directions for future competitions would be to include high-frequency data as well as considerably longer time series. These would offer a new set of challenges but may be especially valuable in forecasting performance comparisons of hybrid and combination methods.

My first exposure to forecasting was through Makridakis and Wheelwright's visionary book (1977). I remember how intrigued I was with their emphasis on the fundamental role of judgment in forecasting and on how judgment is infused throughout the steps in the forecasting process. As a researcher of behavioural dynamics in forecasting, I look forward to seeing this pivotal outlook reflected in the form of clear and structured opportunities for entering forecaster's judgment and expertise in future competitions. Noting that the M2 competition (Makridakis et al., 1993) did investigate this aspect of forecasting to some degree (where participants could seek additional information to incorporate judgment into budget and macro-economic forecasts across 29 series), I believe it would be very timely to return to this emphasis and include a systematized judgmental competition cutting across multiple domains.

Current competition is confined to time series with anonymised variables, thus concealing any contextual information and only allowing judgment to be incorporated through method selection. While selecting a particular technique (or a combination of

techniques) constitutes one of the critical stages of sourcing judgment into the prediction process, other platforms for directly including judgmental forecasts and judgmental adjustments to model forecasts would prove useful in enhancing our understanding of how to improve forecasting performance. This would also enable learning/feedback effects to be investigated via providing outcome and performance feedback to competitors. Examining how such 'reality checks' actually inform expressions of uncertainty (as reflected in probabilistic forecasts, prediction intervals and forecast distributions) would contribute towards the *Competition*'s goals of learning what lies behind the forecasts.

At a time when forecasting is bombarded with criticism and when trust in forecasts appears to be dubious (Goodwin, 2017; Onkal, 2016), Spyros Makridakis continues to empower and inspire us with his innovative portfolio of *M-Competitions* as we look forward to many more.

References:

Goodwin, P. (2017) Forewarned: A Sceptic's Guide to Prediction. London: Biteback Publications.

Hyndman, R. (2018). A brief history of time series forecasting competitions, *Hyndsight blog*, https://robjhyndman.com/hyndsight/forecasting-competitions/, last accessed 10Feb2019.

Makridakis, S., C. Chatfield, M. Hibon, M. Lawrence, T. Mills, K. Ord and L.F. Simmons (1993). The M2-competition: A real-time judgmentally based forecasting study. International Journal of Forecasting, 9(1), 5-22.

Makridakis, S., E. Spiliotis and V. Assimakopoulos (2018a). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting,* https://doi.org/10.1016/j.ijforecast.2018.06.001

Makridakis, S., E. Spiliotis and V. Assimakopoulos (2018b). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE* 13(3): e0194889. https://doi.org/10.1371/journal.pone.0194889

Makridakis, S. and S.C. Wheelwright (1977). *Forecasting: Methods and Applications*, Santa Barbara: Wiley-Hamilton.

Önkal, D. (2016). Trust in Forecasting. Paper presented at the *36th International Symposium on Forecasting*, Santander, Spain.